Asia-Pacific Journal of Atmospheric Sciences
https://doi.org/10.1007/s13143-022-00291-4

Online ISSN 1976-7951
Print ISSN 1976-7633

ORIGINAL ARTICLE

Forecasting $PM_{2.5}$ Concentration in India Using a Cluster Based Hybrid Graph Neural Network Approach

Pavan Sai Santhosh Ejurothu[1] · **Subhojit Mandal**[1] · **Mainak Thakur**[1]

Received: 26 April 2022 / Revised: 29 July 2022 / Accepted: 16 August 2022
© The Author(s) under exclusive licence to Korean Meteorological Society and Springer Nature B.V. 2022

Abstract

Air pollution modeling and forecasting over a national level scale for a country as large as India is a very challenging task due to the large amount of data involved in a limited spatial frequency. Often the air pollution and pollutant dispersion process depend on underlying meteorological conditions. Recently, Graph Neural Networks emerged as an effective deep learning model for discovering spatial patterns for various classification and regression tasks. This study proposes to employ a cluster-based Local Hybrid-Graph Neural Network (HGNN) methodology instead of using a single global Graph Neural Network for monitoring station-wise multi-step $PM_{2.5}$ concentration forecasting across India's states. This methodology respects sudden changes in $PM_{2.5}$ concentration due to the local meteorological variations. However, the local Hybrid GNN models consist of two parts: a spatio-temporal unit containing a Graph Neural Network layer along with a Gated Recurrent Unit layer to model the influence of wind speed and other meteorological variables on $PM_{2.5}$ concentration. The other part is a station wise feature extraction unit to extract station-wise meteorological feature impact on $PM_{2.5}$ concentration, along with the temporal dependency between historical records. The results from the two units are fused in step-wise manner for multi-step $PM_{2.5}$ forecasting. The proposed methodology was used to develop separate $PM_{2.5}$ concentration forecasting models, +24, +48 and +72 hours ahead. Subsequently, a detailed analysis is carried out to unfold the advantages of the proposed methodology. Results demonstrate the proposed models perform better than the state-of-the-art with significantly lesser computation time.

Keywords Air pollution · $PM_{2.5}$ forecasting · Graph neural network · Cluster based graph neural network · Clustering

1 Introduction

Rapid Industrial development and urbanization have facilitated degradation of environment across the globe. Air pollution is one of the recent severe threats to mankind as it affects natural ecological balance and human health. $PM_{2.5}$(Particulate Matter with a diameter of 2.5 micrometers or less) is a known to be one of the most hazardous air pollutants for overall human health. To be specific, it can actually penetrate through human body cells via air exchange in the lungs (Xing et al. 2016). Notably, many of the world's most polluted cities are found in India. In 2017, 77% of India's population was exposed to yearly population-weighted mean $PM_{2.5}$ levels greater than 40 $\mu g/m^3$, the threshold recommended by India's National Ambient Air Quality Standards (Balakrishnan et al. 2019).

It is obvious that reliable forecasts of pollutant concentration can assist decision-makers in taking appropriate actions to control emissions. In view of this, Indian government has launched a monitoring program called the National Air Quality Monitoring Program (NAMP) under the Central Pollution Control Board to track the air quality (Pant et al. 2019). Although many recent studies have been carried out to forecast the air quality for different regions of India (Tiwari et al. 2021; Lal and Tripathy 2012; Gogikar et al. 2019), very limited investigation has been done on the development of a national scale air pollution monitoring and forecasting system for India. To be more specific,

Responsible Editor: Youn-Seo Koo

✉ Mainak Thakur
mainak.t@iiits.in

Pavan Sai Santhosh Ejurothu
pavansaisantosh.e18@iiits.in

Subhojit Mandal
subhojit.m@iiits.in

[1] Computer Science & Engineering Department, Indian Institute of Information Technology, Sricity, Gnan Marg, 517646, Andhra Pradesh, India

Published online: 29 August 2022

for air pollution forecasting, different modeling approaches are followed in the literature such as physical simulation-based modeling approach (CMAQ Byun and Schere 2006, WRF-Chem Forkel et al. 2012 etc.) and data-driven modeling approach. Furthermore, the data-driven approaches, including modern deep learning methods promise better performance, utility and flexibility over the traditional physics-based simulation models in recent times. (Stern et al. 2008; Xu et al. 2020)

Researchers looked at PM_{10} pollution episodes as a function of synoptic climatology and PM_{10} long-range transport (Grivas et al. 2008). Tai et al. (2010) and Spindler et al. (2010) illustrate the impact of meteorological parameters on long-range travel, secondary particle mass production, and re-emission on PM_{10}, $PM_{2.5}$, and PM_1 concentration and composition. The transportation from the pollution sources and deposition of the pollutants are extremely complex phenomena and are dependent on several atmospheric conditions and geographical factors such as terrain information, U-component of wind (eastward wind), V-component of wind (northward wind), relative humidity (Stirnberg et al. 2021), K-index (a measure of thunderstorm potential of lower atmosphere), 2m Temperature (air temperature at 2 meters above the surface). In addition, the planetary boundary layer of the atmosphere is responsible for thermodynamic perturbation of the atmospheric variables and governs major pollutant characteristics such as deposition, transportation, transformation, and diffusion (Miao et al. 2019). These important weather variables are often considered as part of inputs for data-driven pollutant concentration forecasting methodology.

Several deep learning models such as LSTM (Long Short Term Memory) (Karimian et al. 2019), Bi-directional LSTM (Tong et al. 2019), Attention-based LSTM (Li et al. 2021), GRU (Gated Recurrent Units) (Becerra-Rico et al. 2020), Image based convolutional Neural Network (Chakma et al. 2017; Ma et al. 2018), Diffusion Convolutional Recurrent Neural Network (GC-DCRNN) (Lin et al. 2018), Encoder-decoder based models have been used in monitoring station based pollutant concentration forecasting and weather prediction problem. Although these models became successful in capturing the long or short term temporal dependency, they fail to address spatial dependency between the measurements of the monitoring stations located geographically far or near to each other. To alleviate this issue, neural networks on graph structured data were used for station based $PM_{2.5}$ concentration forecasting purpose, considering monitoring stations as nodes. Correspondingly, Qi et al. (2019) proposed GC-LSTM model, which combines Graph convolutional networks and LSTM networks to model and forecast $PM_{2.5}$ concentration for 76 monitoring stations in China. Convolutional filters (Defferrard et al. 2016) on graph structures are used to learn the local stationary features of the underlying process. But, graph-convolution based models (eg.: GC-LSTM) fail to extract spatio-temporal features efficiently. Wang et al. (2020) developed $PM_{2.5}$-GNN model which uses a combination of GNN module and a spatio-temporal GRU module to forecast $PM_{2.5}$ concentrations over 184 cities in China. Though this method worked well but only a limited part of China was covered for $PM_{2.5}$ predictions in the mentioned study. Furthermore, the suggested model $PM_{2.5}$-GNN (Global-GNN) did not address the stationarity challenges of the underlying random process. As a result, while applying such a model over a broad area such as all over India, training the suggested model on a vast graph structure may not be suitable. Such large graph structures are bound to result in unwanted problems due to the inherent non-stationarity of the random process responsible for pollutant concentrations of significantly large landscape. Galmarini et al. (2018) suggests the use of regional-scale models in an ensembled manner promises smaller long-term error values than the global models. Accordingly, hybrid ensemble approaches can be employed in the modeling methodology, where developing a robust model often depends on the scale of operation the modeler chooses. The efficient scale selection is deeply rooted in the atmosphere modeling. The approach is primarily related to the energy spectrum associated with the scale separation of atmospheric variables (Van der Hoven 1957). The success of regional ensembled modeling approach motivates us to develop a robust $PM_{2.5}$ forecasting model comprises of regional forecasting models, which respects the local meteorological changes efficiently compared to a global forecasting model.

In this work, we propose a Cluster based Local Graph Neural Network (CLGNN) methodology to partition a large graph structure into smaller local clusters in order to melt down the global non-stationary process on large landscape to local stationary ones using a smaller graph structure. The model is developed for forecasting the $PM_{2.5}$ concentrations at monitoring stations all over India. Several clustering methods have been considered to partition the monitoring stations and the CLGNN models were developed for each of them. The proposed methodology suggests to use regional ensemble hybrid graph neural networks For each cluster based CLGNN models, multi-step forecasting capabilities were investigated through the development of +24 hours, +48 hours and +72 hours forecasting models and corresponding detailed analysis was done. The proposed model comes with reduced time complexity and better accuracy than the global model. This paper is organized in the following manner: Section 2 presents the dataset and methodology. The proposed model components have been shown as well. Section 3 presents the experiment settings and results. The discussions related to different methods are also presented here. Section 4 presents the contribution and conclusion of this study.

2 Data and Methodology

2.1 Study Area and Dataset

This study focuses on the $PM_{2.5}$ concentration forecasting for India ($8°4' \sim 37°6'$N and $68°7' \sim 97°25'$E). The increasing anthropogenic emissions (burning of fossil fuel from households (Rao et al. 2021), industrial emissions, bio-mass burning) is causing the rapid increment of particulate matters in the atmosphere. Particularly, Indo-Gangetic Plain (IGP) in the northern part of India is one of the world's most polluted areas. As a consequence, the capital city, New Delhi, is facing high pollution episodes every year during winter season due to increased traffic conditions and domestic bio-mass fuel consumption.

We employ two types of data: firstly, half hourly air quality $PM_{2.5}$ data from $N_{ms}(= 258)$ Ground Monitoring Stations (GMS) for continuous ambient air quality monitoring, covering different regions of India from CPCBCCR (Central Control Room For Air Quality Management) (CPCB Central Pollution Control Board 2001). The other type of data used is meteorological variables data (wind speed, wind direction, precipitation, planetary boundary layer height, K-index, surface pressure, u,v-component of wind speed, relative humidity, 2-m temperature). We make use of climate reanalysis ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Copernicus Climate Change Service 2017). The GMS locations are shown in Fig. 1. The $PM_{2.5}$ concentration for all the GMSs are converted into 3-hour interval dataset to match with meteorological data. Let for each N_{ms} no. of GMSs, $PM_{2.5}$ and meteorological variables' multivariate time series data with l no. of historical records taken, where each historical length represents 3 hours of a time interval. The aim of the study is to predict for each N_{ms} no. of GMSs, $PM_{2.5}$ concentrations for next k future timestamps (Each future timestamp resolution is 3 hours). Let g be the function, that represents the short term forecasting model all over India. Then g is given by

$$g\left(X^{t}_{PM_{2.5}}, X^{t}_{Met}\right) = \left[x^{t+1}_{PM_{2.5}}, \dots, x^{t+k}_{PM_{2.5}}\right] \tag{1}$$

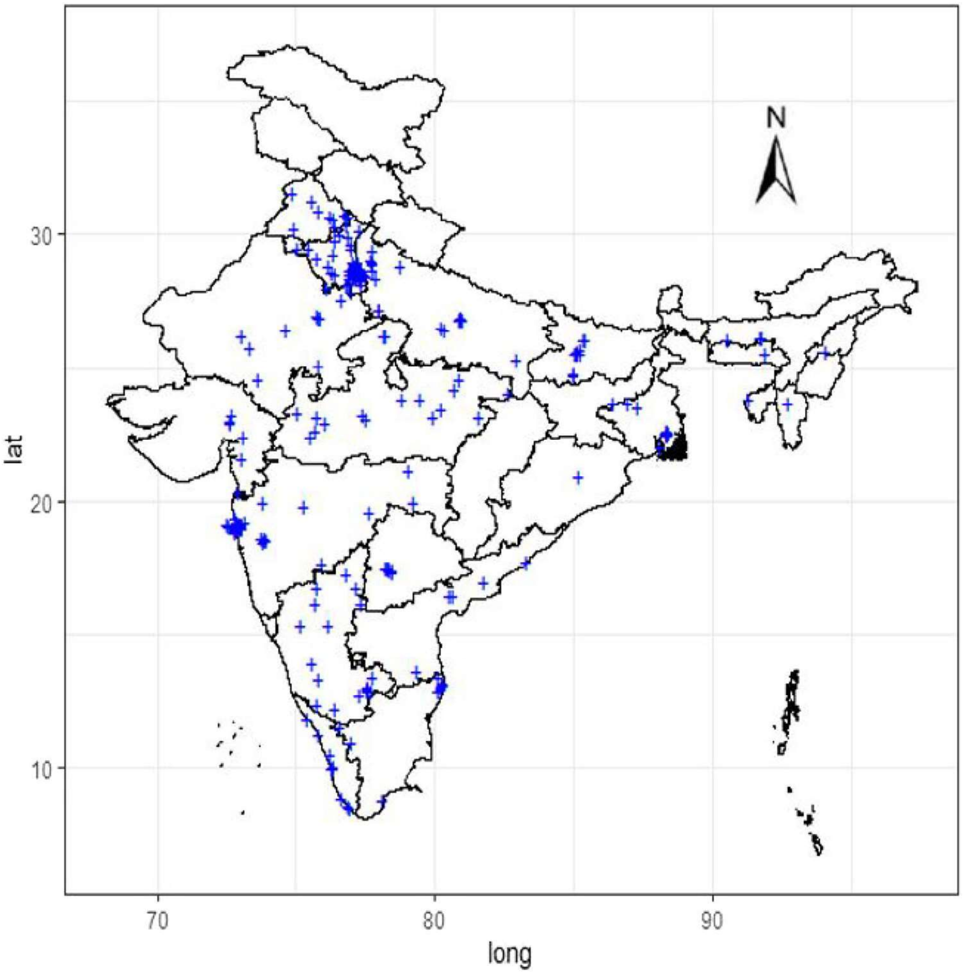


Fig. 1 Ground monitoring stations in India

$X^t_{PM_{2.5}} \in R^{N_{ms}\times l}$, $X^t_{Met} \in R^{N_{ms}\times l\times d}$, d and l is the number of meteorological features and historical length respectively.

2.2 Methodology

The proposed methodology employed for India's air pollution framework is presented in Fig. 2. This study proposes a cluster-based method to approximate Eq. 1 by local hybrid graph neural networks on disjoint clusters of GMSs. Accordingly, several clustering techniques are explored in this study for dividing the large graph structure of the monitoring stations into smaller disjoint ones. The following clustering techniques were explored: Weighted k-means, Spectral clustering and Hierarchical clustering. Weighted k-means clustering algorithm groups the monitoring stations considering the $PM_{2.5}$ concentrations as weights along with the geo-coordinates of the stations (Huang et al. 2005). The spectral clustering method uses the spectral properties of an affine matrix (A), constructed from the distances between monitoring stations (Ng et al. 2001). The hierarchical clustering creates a (often binary) tree out of the data. Individual data points are represented by the leaves, but the root is a single cluster that holds all of the data. Several studies use different statistical measures for determining optimal number if clusters e.g.: Silhouette score (Rousseeuw 1987), Dunn index, within-cluster sum of squares (WCSS). The optimal number of clusters is determined using Silhouette score (Huang et al. 2005; Beddows et al. 2009) and Davis-Bouldin score (Davies and Bouldin 1979) in this study. The silhouette value is a measure of how similar an object is to its cluster (cohesion). A high value implies that the object is well matched to its own cluster and poorly matched to surrounding clusters, and a low value suggests that, the object is badly matched to surrounding clusters. The optimal number of clusters were decided for each clustering method. Let $C = \sqcup_{i=1}^{k} C_i$, C is the set of all GMSs, C_i is the i^{th} cluster. For each cluster C_i, the local hybrid-GNN model is employed.

Let $X_t = \left(X^t_{PM_{2.5}}, X^t_{Met}\right)$ is a sample for the local hybrid GNN model, where $X^t_{PM_{2.5}}\left(=\left[x^{t-l}_{PM_{2.5}}, \ldots, x^{t}_{PM_{2.5}}\right]\right)$ represents the $PM_{2.5}$ concentration and X^t_{Met} represents meteorological features for all GMSs with historical values. Notably, the value l represents the no. of historical time stamp information during each time forecasting. This hybrid GNN model approximates a function g_W, where $g_W(X_t) = \left[x^{t+1}_{PM_{2.5}}, \ldots, x^{t+k}_{PM_{2.5}}\right]$. g_W learns the complex relationship through the corresponding weight matrices involved in the model architecture. For each cluster C_i, let the function g_{W_i} approximate the local hybrid GNN model. This study focuses on the modeling of the underlying global function g which approximates the short term forecasting model through local g_{W_i} functions for corresponding the cluster of GMSs. The function g can be represented as: $\oplus_{i=1}^{k} g_{W_i}$. For Weighted k-means, Spectral and Hierarchical clusters, we denote the ensemble set of local models

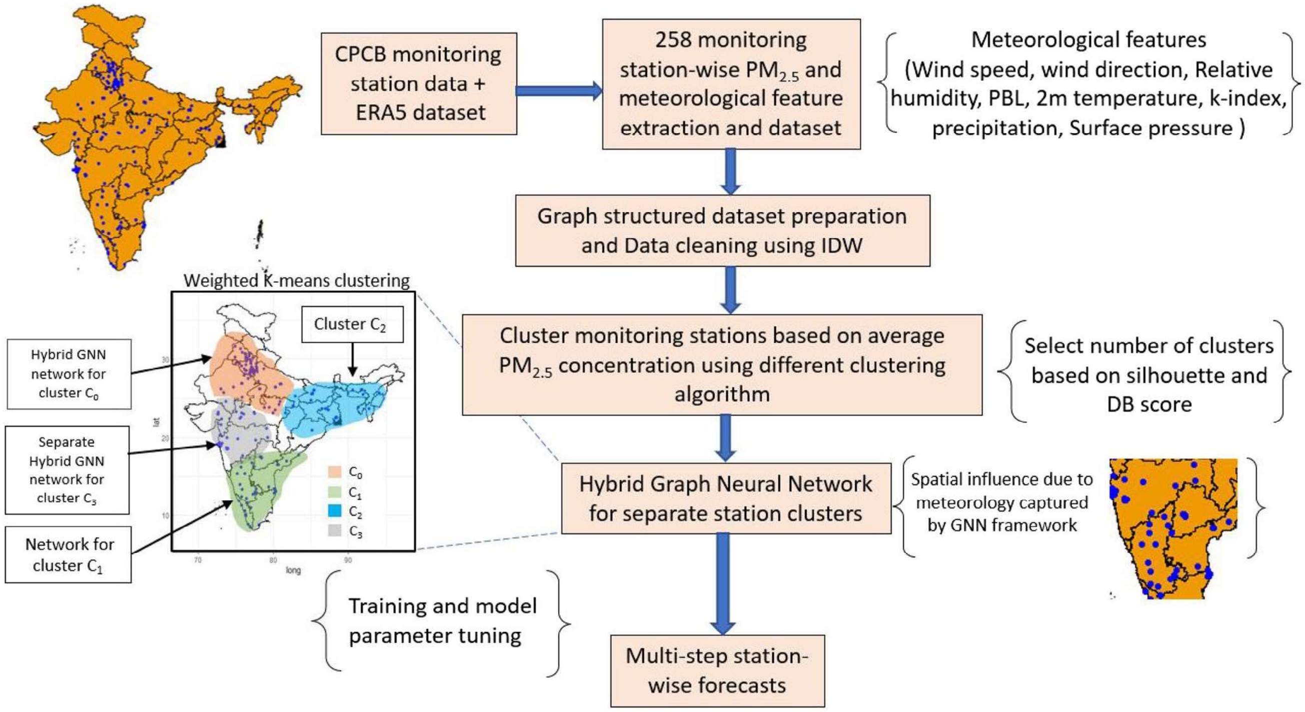

CPCB monitoring station data + ERA5 dataset
258 monitoring station-wise $PM_{2.5}$ and meteorological feature extraction and dataset
Meteorological features (Wind speed, wind direction, Relative humidity, PBL, 2m temperature, k-index, precipitation, Surface pressure)
Graph structured dataset preparation and Data cleaning using IDW
Weighted K-means clustering
Cluster C_2
Hybrid GNN network for cluster C_0
Separate Hybrid GNN network for cluster C_3
Network for cluster C_1
C_0
C_1
C_2
C_3
30
20
10
lat
70
80
90
long
Cluster monitoring stations based on average $PM_{2.5}$ concentration using different clustering algorithm
Select number of clusters based on silhouette and DB score
Hybrid Graph Neural Network for separate station clusters
Spatial influence due to meteorology captured by GNN framework
Training and model parameter tuning
Multi-step station-wise forecasts

Fig. 2 Methodology framework

as WCLGNN, SCLGNN and HCLGNN model respectively. The computational time for the clustered GNN model is computed as the sum of the times taken for cluster-wise local models.

2.2.1 Hybrid Graph Neural Network Model (HGNN)

This study proposes to use local Hybrid Graph neural network models (HGNN) on different clusters for capturing local spatial variability efficiently. This hybrid model processes the data through two different units. One for spatio-temporal feature extraction (Spatio-temporal unit) and another unit for station-wise feature extraction. Spatio-temporal features for a monitoring station relate to the nearest neighbor node's meteorological variables and $PM_{2.5}$ concentration influence factors. Also the mixture of these influence factors along with the local stations' local meteorological conditions interact with each other to extract relevant features for each node in the spatio-temporal feature extraction unit. On the other hand, station-wise feature extraction relates to the local embedding of $PM_{2.5}$ and meteorological features for individual stations. Both units are employed for $PM_{2.5}$ forecasting to utilize the neighbor influence based patterns as well as the local station $PM_{2.5}$-meteorological interaction patterns. The outputs from both the units are concatenated and sent through a final MLP layer to forecast on the subsequent timestamps. The architecture of the network is shown in Fig. 3.

Spatio-Temporal Unit (STU) The spatio-temporal unit (STU) of the model consists of a graph neural network module and 2 spatio-temporal GRU (Gated Recurrent networks), Multi-layer Perceptron (MLP) layer (Rumelhart et al. 1985). GRU cells were proposed by Chung et al. (2014). The architecture of the spatio-temporal unit closely follows Wang et al. (2020). Wind speed and direction play an important role when the air pollutant diffusion and transportation phenomenon is considered (Hu et al. 2014; Miao et al. 2019). To capture the air pollutant diffusion and transportation information between the monitoring stations the graph neural network module uses an advection coefficient (proposed by Wang et al. 2020) feature for each GMS, which is given by $S_{ji}^{t} = ReLU(\|v\|cos(\theta)/D)$, where v is the wind speed of the source node j, D is the distance between j and sink node i. This feature acts as an edge feature. Notably, the source node and sink nodes represents as the node where pollutant transports from source GMS to sink GMS. The node-wise meteorological feature and $PM_{2.5}$ values interact with the edge feature through an edge MLP layer, which uses a series of MLP layers and sigmoid activation functions. Specifically, the following features: source node meteorological feature and $PM_{2.5}$, sink node same features, edge attribute features, along with the edge weight (advection coefficient) are fused through edge MLP layer during each time stamp.

$$\xi_i^t = \left[X_i^{t-1}, P_i^{t-1}\right], \forall i \in V \tag{2}$$

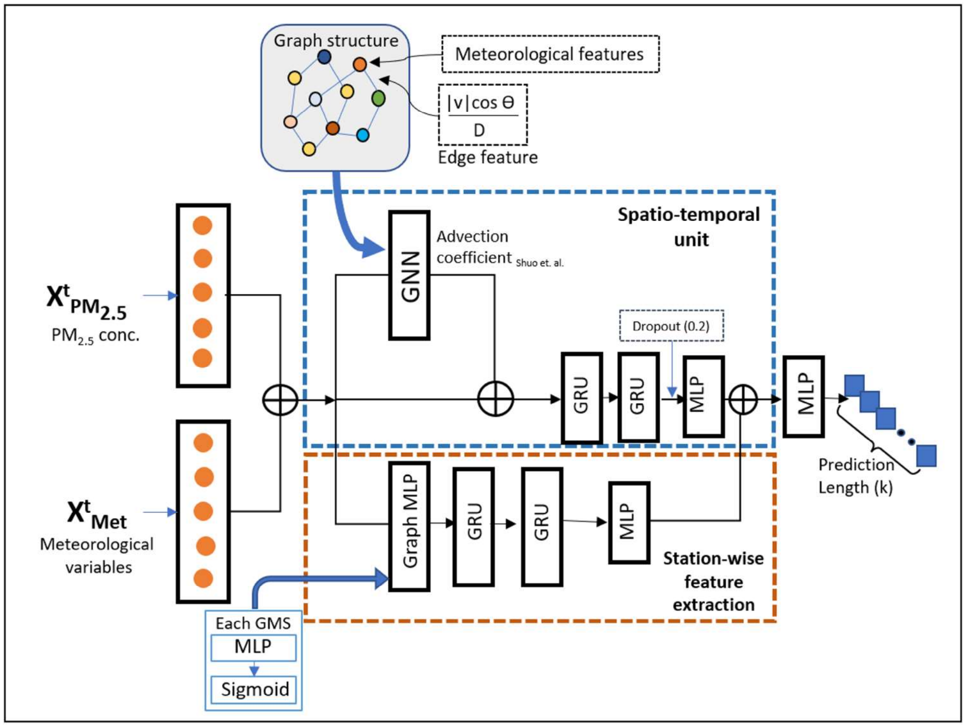


Fig. 3 Hybrid graph neural network

$$e_{ji}^{t} = \Psi\left(\left[\xi_{j}^{t}, \xi_{i}^{t}, S_{ji}^{t}\right]\right), \forall (j, i) \in E \tag{3}$$

$$\zeta_{i}^{t} = \Phi\left(e_{ji}^{t}, e_{ij}^{t}\right)_{j \epsilon N(i)}, \forall i \in V \tag{4}$$

The edge attribute X_{i}^{t-1} is the $\mathrm{PM}_{2.5}$, P_{i}^{t-1} is the meteorological feature, these features act as node-wise feature at previous time-stamp. Ψ, Φ is the edge MLP layer and node MLP layer respectively. After fusing node-based information with advection coefficient feature through edge MLP layer. The node-wise feature is extracted from the output of the edge MLP layer and passed through a node MLP layer along with sigmoid activation function. This spatio-temporal feature is used in the Gated Recurrent Unit networks, which consist of MLP layers and GRU cell. Also, the spatio-temporal GRU networks forecasts multi-step features for each monitoring stations. These multi-step features are used for multi-step $\mathrm{PM}_{2.5}$ forecasting along with features from the station-wise feature extraction unit.

The nodes of the graph neural network follow a message passing scheme to encode the neighboring nodes' meteorological influence on the $\mathrm{PM}_{2.5}$ propagation. On the top of that, the temporal diffusion of the $\mathrm{PM}_{2.5}$ is well captured by spatio-temporal GRU networks. At each time step, the GRU network takes the input as the combination of the node-wise $\mathrm{PM}_{2.5}$, meteorological feature, and also, the GNN extracted transportation-related features. As the node-wise data is considered in the same GRU layer, the spatio-temporal GRU network learns related spatial transport features between the monitoring stations, parallelly learning temporal dependency.

Station-Wise Feature Extraction Unit (SWFU) The station-wise feature extraction module (SWFU) is used for extracting relevant features from the local embedding of the station-wise $\mathrm{PM}_{2.5}$ and meteorological features. This unit contains graph MLP layer, GRU networks, and MLP layers. The graph MLP layer extracts the station-wise independent $\mathrm{PM}_{2.5}$ concentrations variation based feature, and let them pass through GRU networks for modeling the temporal dependency out of the feature. Notably, the graph MLP uses multi-layer perceptron layer along with sigmoid non-linear activation function for extracting station-wise $\mathrm{PM}_{2.5}$ and meteorological fluctuation related patterns. Precisely, this layer embeds the station-wise $\mathrm{PM}_{2.5}$ and meteorological features for each monitoring stations and utilizes the non-graphical structure of the data to provide noise-free robust features for subsequent GRU networks. The process in the graph MLP layer is described by the following Eq. 5.

$$z_i = \sigma(MLP(F_i)), \forall i \in V \tag{5}$$

The MLP(.) function is the multilayer perceptron, F_i is the concatenated $\mathrm{PM}_{2.5}$ and meteorological features at the i^{th} node. The $\sigma(.)$ function is the sigmoid activation function (Goodfellow et al. 2016). Also, the MLP(.) function is shared across the nodes. In an important note, Hu et al. (2021) recommended using graph MLP-based designs rather than graph neural network designs for classification-based tasks while dealing with large scale graph data and corrupted adjacency information. In our situation, the redundant edge features are included in the STU unit's GNN layer and forwarded to the GRU networks, which reduces the unit's forecasting capability. On the other hand, graph MLP provides the relevant features over the far distant, irrelevant edge information related to the features of the advection coefficient in the STU unit. In fact, the graph MLP layer uses the non-graphical structure of the data to enable the overall modeling to employ a combination of non-graph structured and graph structured methodology for $\mathrm{PM}_{2.5}$ forecasting.

The extracted station-wise patterns from graph MLP layer are sent to 2 GRU networks for modeling the individual station-based temporal dependency. In the final layer of the SWFU unit, an MLP layer is used to project the GRU extracted feature into an intermediate feature, regarded as the output of the SWFU unit. Further, the output from the SWFU unit is sent to concatenated with the output from the STU unit.

2.2.2 Prediction Module

The last module of the model is an MLP layer. For each prediction time stamp, the module uses the concatenated features from STU and SWFU and provides the $\mathrm{PM}_{2.5}$ forecasts as output. Let h_{st}^{t} output from the spatio-temporal unit and h_{swf}^{t} output from station-wise module. $g_{pred}\left(h_{st}^{t} \oplus h_{swf}^{t}\right) = \left[x_{PM_{2.5}}^{t+1}, \dots, x_{PM_{2.5}}^{t+k}\right], h_{st}^{t}, h_{swf}^{t} \in R^{N_{ms} \times k}$.

2.2.3 Comparison Metrics

To evaluate the performance of the model, we use the following evaluation metrics: $RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2\right)}$, $MAE = \left(\frac{1}{n}\sum_{i=1}^{n}\|Y_i - \hat{Y}_i\|\right)$, $R^2 = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2)}{\sum_i (Y_i - \bar{Y})^2)}$ where Y_i, $\hat{Y}_i$ are the i^{th} ground truth and prediction respectively, $\bar{Y}$ is the mean of the predictions and n is the total number of data samples observed. The coefficient of determination (R^2) metric provides a measurement of proportion of the actual variation that is explained by the model. The epoch-wise train loss and test losses are calculated to evaluate the models' generalization ability based on MAE, RMSE, R^2 between the actual and predicted values. To check a few other important performance metrics, we binarize the prediction and ground truth to 0/1 matrices based on a cut-off. This cut-off value denoting pollutants concentrations exceed the demarcation point (suggested by NAAQS in 2009) 60 $\mu g/m^3$ or not. Generally, the skill of the models are presented by following metrics CSI (Critical Success Index), POD (Probability Of Detection), and FAR (False Alarm Rate) (Kang et al. 2007). Then, to predict CSI, POD, FAR; we calculate hits (prediction = 1, truth =1), misses (prediction =0, truth =1), and false alarms (prediction = 1, truth =0) to evaluate the performance of model. $CSI = \frac{hits}{hits+misses+flasealarms}$, $POD = \frac{hits}{hits+misses}$, $FAR = \frac{hits}{hits+falsealarms}$. POD represents the model's capability to forecast extreme events. CSI provides a score that indicates how correctly extreme events are forecasted, considering false alarm rates and misses. FAR is concerns with the false alarm the model provides. CSI, FAR and POD values lie between 0 to 1. The ideal situation indicates CSI and POD provide value 1. Similarly, FAR provides 0 for an ideal situation.

3 Results and Discussion

3.1 Experimental Setup

The dataset consists of 258 GMS's 3-hourly $PM_{2.5}$ and meteorological data which spans two years and three months (01-01-2018 to 31-03-2020). We divided the whole dataset into two parts for training (01-01-2018 to 31-12-2019) and testing (01-01-2020 to 31-03-2020). For each variable, Inverse Distance Weighting (IDW) interpolation method (Shepard 1968) has been used for spatial-gap filling. Each of the samples consists of 3 parts, $\left(X^t_{PM_{2.5}}, X^t_{Met}, \left[x^{t+1}_{PM_{2.5}}, \ldots, x^{t+k}_{PM_{2.5}}\right]\right)$, $x^{t+k}_{PM_{2.5}}$ is the actual $PM_{2.5}$ value. Notably, $x^{t+k}_{PM_{2.5}}$ works as label while training the model. The value of l in $X^t_{PM_{2.5}}\left(=\left[x^{t-l}_{PM_{2.5}}, \ldots, x^{t}_{PM_{2.5}}\right]\right)$ is chosen as 8, as the increment of the value attracts high training time of the model.

Along with the meteorological features, the following features were used in this study: day of the month, day of the week. All the features and $PM_{2.5}$ data were standardized using the mean and variance of training data. To investigate the variation of cluster-wise $PM_{2.5}$ trends, the monthly $PM_{2.5}$ average is provided in Fig. 4 for weighted k-means clustering. It shows that for GMSs in the same clusters, the trends are similar, and in the different clusters, the trends are varying. The boxplot for $PM_{2.5}$ concentrations for GMSs in different clusters are shown in Fig. 5. This shows the data variability across the clusters. The exploratory data analysis is shown in Table 1. The average $PM_{2.5}$ all over India within the study period 71.2 $\mu g/m^3$. The boundary layer data shows to have high standard deviation.

The station-wise $PM_{2.5}$ average is used as input for the clustering algorithms. For weighted k-means clustering, as the silhouette score for 4 clusters are high and davis-bouldin score is less (Fig. 6), we perform forecasting using 4 local HGNN models for 4 clusters (as shown in Fig. 7). Similarly, for Spectral clustering method, 4 local HGNN and for Hierarchical clustering method 2 HGNN models were developed. The Mean Squared Error loss function is used for the optimization of the local HGNNs. All the models are trained by Adam optimizer along with a learning rate of 0.01, and batch size 8 for 50 epochs for each experiment. Weighting decay of 0.0005 and gradient accumulation (Wu et al. 2018) technique was used to avoid overfitting of the model. Learning rate scheduler (Senior et al. 2013) was employed for decaying the learning rate of each parameter group by gamma value 0.1 and step size 15. The hidden size of all the GRU cells is 128 for all the GRU layers. For WCLGNN, HCLGNN, SCLGNN models, the next 24, 48, 72 hours ($k = \{8, 16, 24\}$ respectively) forecasting capabilities were explored.

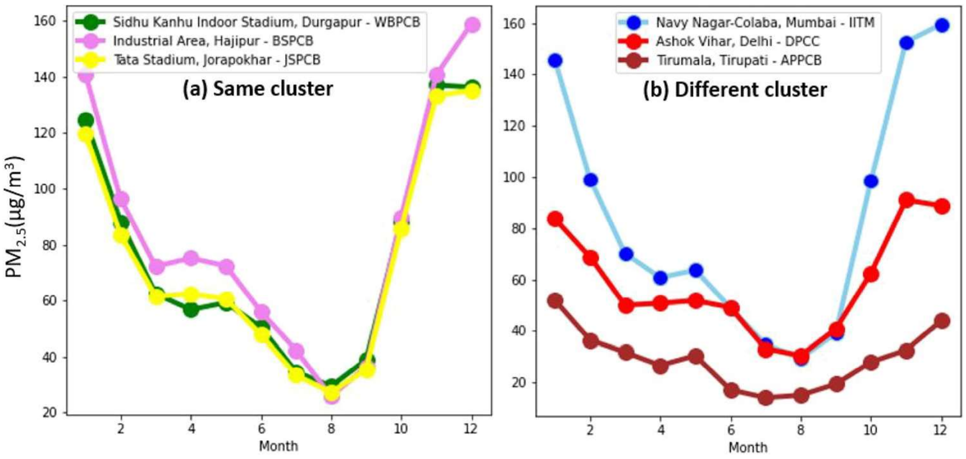


Fig. 4 Monthly trends of monitoring stations for same clusters and different clusters

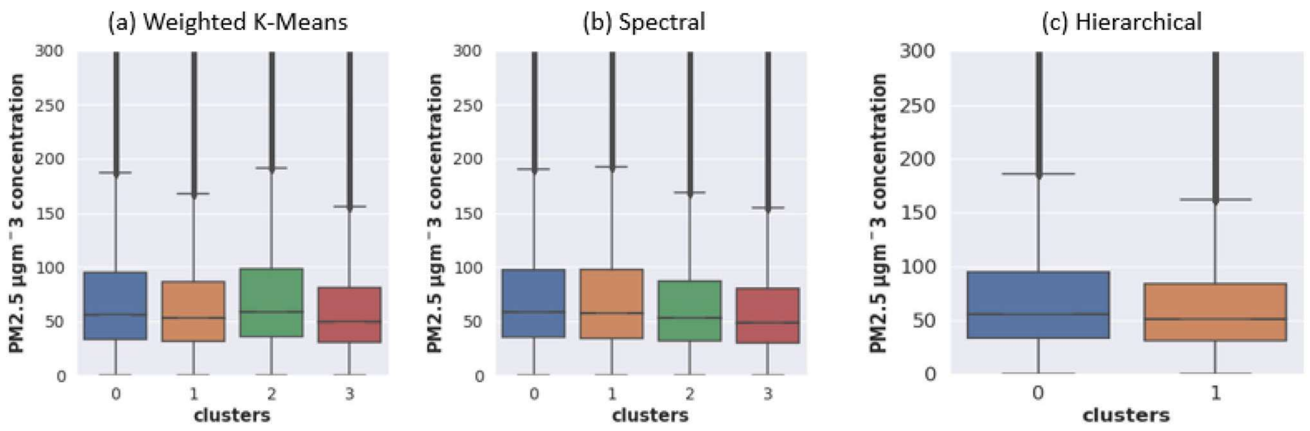
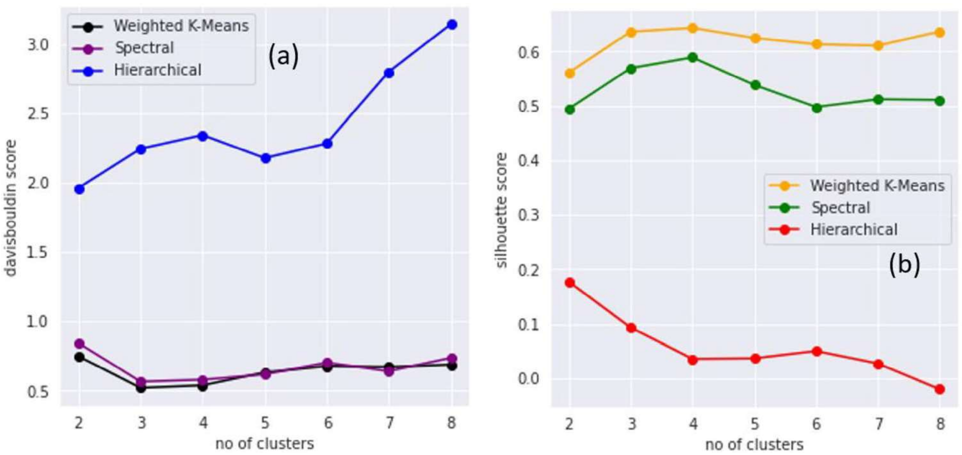


Fig. 5 Box plots for different clusters based on (**a**) Weighted K-means (**b**) Spectral (**c**) Hierarchical clustering methods

Fig. 6 Graphs for finding optimal number of clusters (**a**) Davis Bouldin score (**b**) Silhouette score for each clustering algorithm



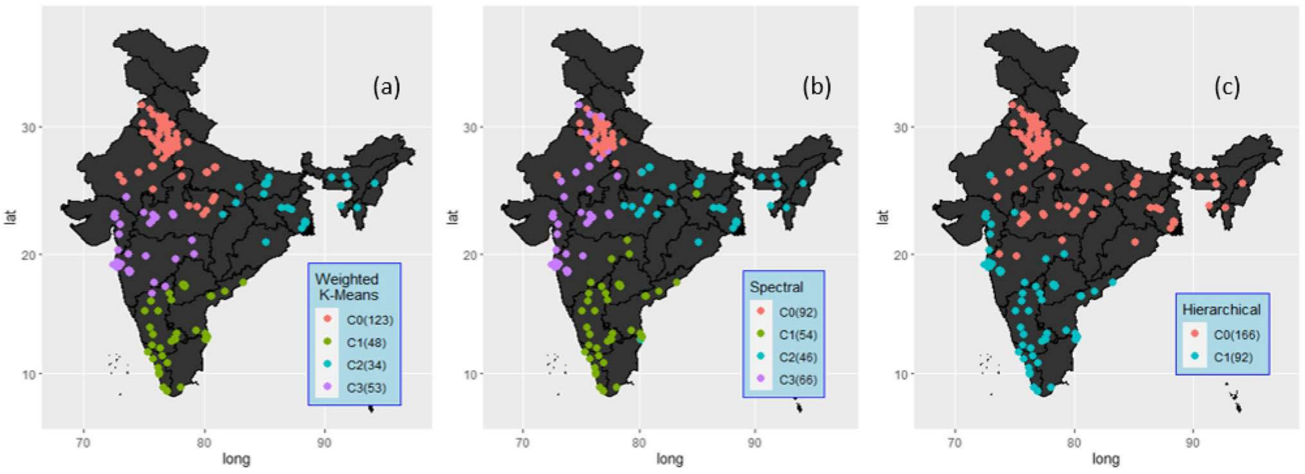

(a)
(b)
(c)
lat
long
30
20
10
70
80
90
Weighted K-Means
C0(123)
C1(48)
C2(34)
C3(53)
Spectral
C0(92)
C1(54)
C2(46)
C3(66)
Hierarchical
C0(166)
C1(92)

Fig. 7 GMS clusters for different clustering algorithms (**a**) Weighted k-means (**b**) Spectral (**c**) Hierarchical clustering method

Table 1 Exploratory data analysis

Atributes (unit)	Mean	Standard Deviation	Minimum	Median	Maximum
PM$_{2.5}$ ($\mu g/m^3$)	71.2	62.6	0.01	54	1000
Wind speed (m/s)	1.37	0.99	0.01	1.19	50
Wind Direction (deg)	106.2	80.3	0.01	77.1	360
Precipitation (m)	0.0002	0.0007	0.0000	0.0000	0.0372
Boundary layer (m)	636.3	603.6	8.6	568.8	6293.8
K-index (K)	14.6	22.4	−0.01	22.6	46.8
Surface pressure (Pa)	93908.01	15226.1	52170.0	100193.0	102245.9
U-component (m/s)	1.10	4.87	−36.25	0.37	39.84
V-component (m/s)	0.95	4.16	−39.01	0.51	40.58
Relative humidity (%)	73.3	24.0	1.79	82.3	100.0
2-m temperature (K)	279.04	19.52	228.89	299.44	320.13

Table 2 Comparision between different baseline models for k=8

Metric	Train loss	Test loss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR
MLP	0.7197	0.3786	31.99	28.13	0.5608	0.8853	0.3952
LSTM	0.4105	0.2056	22.59	19.26	0.7078	0.8821	0.2182
GRU	0.5051	0.1871	20.55	17.11	0.7093	0.8192	0.1591
GC-LSTM	0.4414	0.1910	21.07	17.57	0.7107	0.8389	0.1770
nodesFC-GRU	0.3884	0.1821	20.76	17.48	0.7197	0.8770	0.1996
PM$_{2.5}$-GNN (Global-GNN)	0.3840	0.1723	19.83	16.61	0.7268	0.8637	0.1790

Table 3 Comparision between different models for k=16

Metric	Train loss	Test loss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR
MLP	0.7556	0.4195	34.25	29.47	0.5347	0.8918	0.4282
LSTM	0.4471	0.2308	24.37	20.12	0.6886	0.9011	0.2550
GRU	0.4541	0.2220	23.56	19.26	0.6940	0.8506	0.2097
GC-LSTM	0.4327	0.2221	23.68	19.43	0.6950	0.8770	0.2299
nodesFC-GRU	0.4273	0.2091	23.09	18.92	0.7015	0.8665	0.2135
PM$_{2.5}$-GNN (Global-GNN)	0.4265	0.1945	21.81	17.64	0.7087	0.8394	0.1802

Table 4 Comparision between different models for k=24

Metric	Train loss	Test loss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR
MLP	0.7514	0.4276	35.72	30.40	0.5358	0.9189	0.4376
LSTM	0.7593	0.4315	34.78	29.28	0.5274	0.9507	0.4578
GRU	0.4921	0.2281	24.42	19.67	0.6850	0.8386	0.2109
GC-LSTM	0.4642	0.2474	25.78	20.91	0.6751	0.8634	0.2442
nodesFC-GRU	0.4701	0.2309	25.03	20.40	0.6826	0.8874	0.2527
PM$_{2.5}$-GNN (Global-GNN)	0.4278	0.2084	23.40	18.76	0.6942	0.8680	0.2238

Table 5 Comparision of all models for different predicted lengths

Predicted length (k)	Cluster	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	R^2	Time Taken (min)
24	WCLGNN	22.41	17.93	0.611	1043
	SCLGNN	22.55	18.11	0.603	905
	HCLGNN	22.23	17.64	0.607	1411
	Global-GNN	23.40	18.76	0.562	2613
16	WCLGNN	21.30	17.37	0.637	708
	SCLGNN	21.46	17.47	0.635	466
	HCLGNN	21.39	17.29	0.628	1030
	Global-GNN	21.81	17.64	0.600	1957
8	WCLGNN	19.12	15.91	0.687	419
	SCLGNN	19.40	16.26	0.683	297
	HCLGNN	19.31	16.09	0.682	682
	Global-GNN	19.83	16.61	0.660	1310

3.2 Base Line results

The model performance comparison between other deep learning models and baseline model (for sample type of prediction length 8, 16, 24 with historical length 8) is shown in Tables 2, 3 and 4. The following baseline models have been implemented to check the superiority of the proposed methodology: MLP, LSTM, GRU, GC-LSTM, nodesFC-GRU model (proposed by Wang et al. 2020), $PM_{2.5}$-GNN. Among these models, GC-LSTM model follows a graph structure of the data and applies convolution operation on the node features. As the proposed models follow the graph structure, GC-LSTM is the first proposed model in this context. It can be observed, the $PM_{2.5}$-GNN model all over 258 monitoring stations show superiority in terms of accuracy over the testdata for $k = 8, 16, 24$. In our methodology, we call $PM_{2.5}$-GNN as Global-GNN afterwards in order to avoid conflict. Also for different prediction length $k = 8(+24hour), 16(+48hours), 24(+72hours)$ the developed

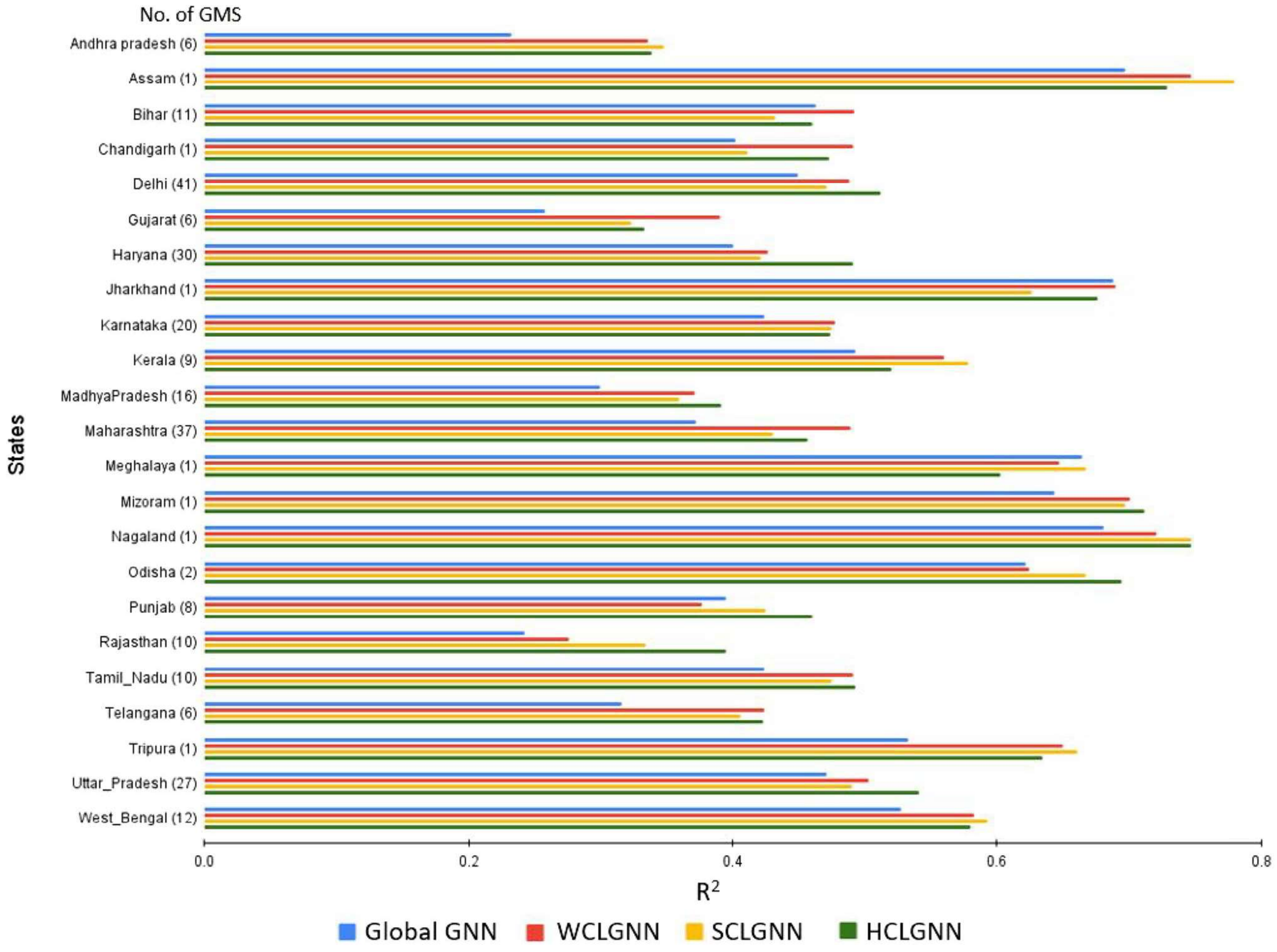


Fig. 8 Comparison of proposed model's R^2 score on state-wise GMS for k=8

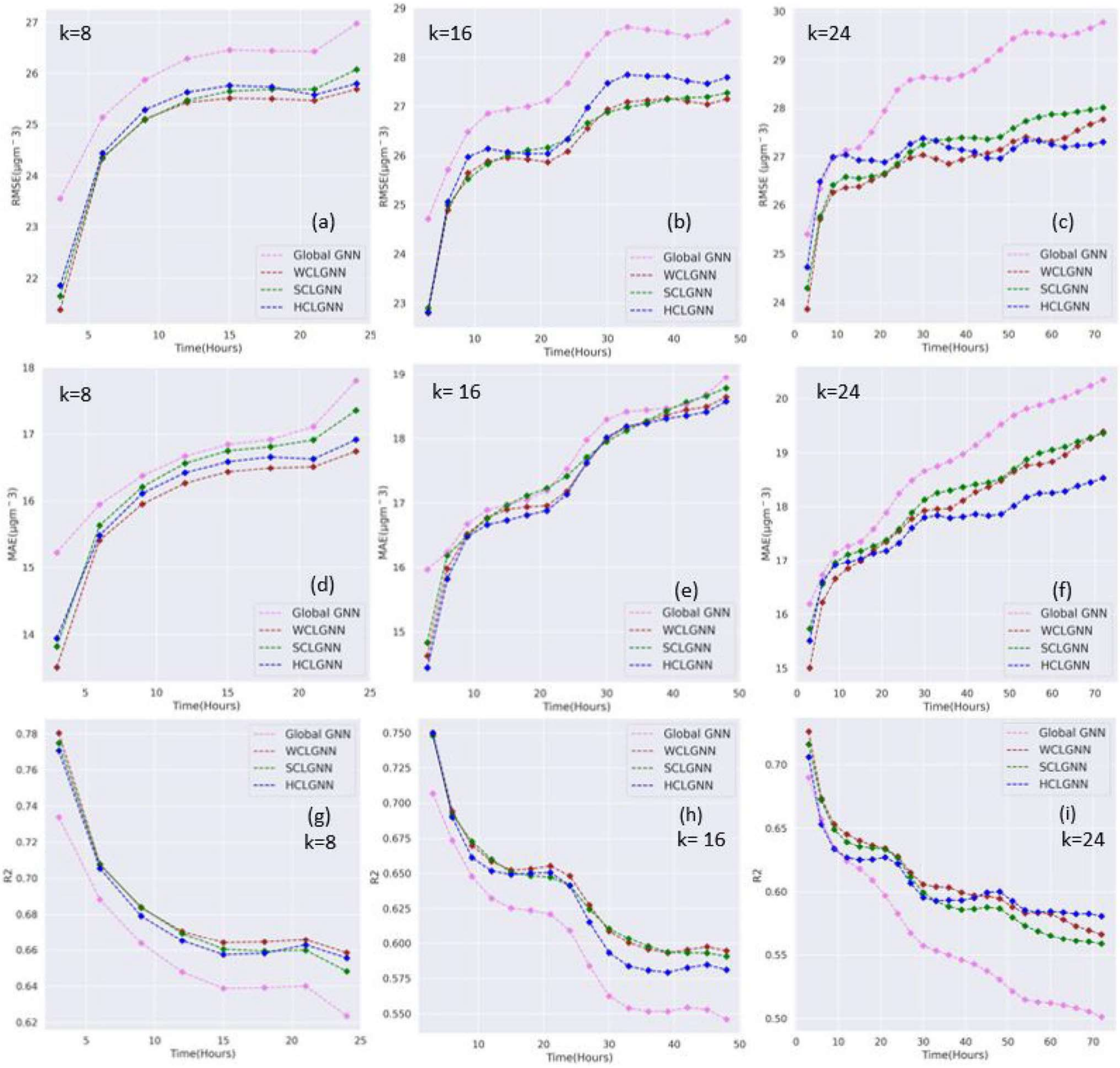


Fig. 9 Prediction step-wise model performance: (**a**), (**d**), (**g**) are 24-hour (k=8) model metrics RMSE, MAE and R^2 respectively; (**b**), (**e**), (**h**) represents the same metrics for 48-hour (k=16) and (**c**), (**f**), (**i**) refers RMSE, MAE and R^2 for 72-hour (k=24) model

$PM_{2.5}$-GNN model shows highest Critical Success Index (CSI) and minimum False Alarm Rate (FAR) which proves the better skill of Global-GNN model.

The comparative results of WCLGNN, SCLGNN and HCLGNN with respect to the global model ($PM_{2.5}$-GNN) is shown in Table 5 for different prediction length (k). WCLGNN model (for $k = 24$) achieves the highest R^2 score 0.61 compared to Global-GNN with R^2 value 0.56. The training time (for $k = 24$) (presented in Table 5) is reduced by $60\%, 65\%, 46\%$ for WCLGNN, SCLGNN and HCLGNN model respectively compared to the Global-GNN model which is a significant reduction.

3.3 Model Performance

The state-wise analysis for model prediction capabilities is shown in Fig. 8. The model scores comparatively poor R^2 value on the test data for GMSs in Gujarat, Madhya Pradesh, Rajasthan, Andhra Pradesh (A.P.), Telengana. Among them GMSs in Andhra Pradesh, Telangana reports low average

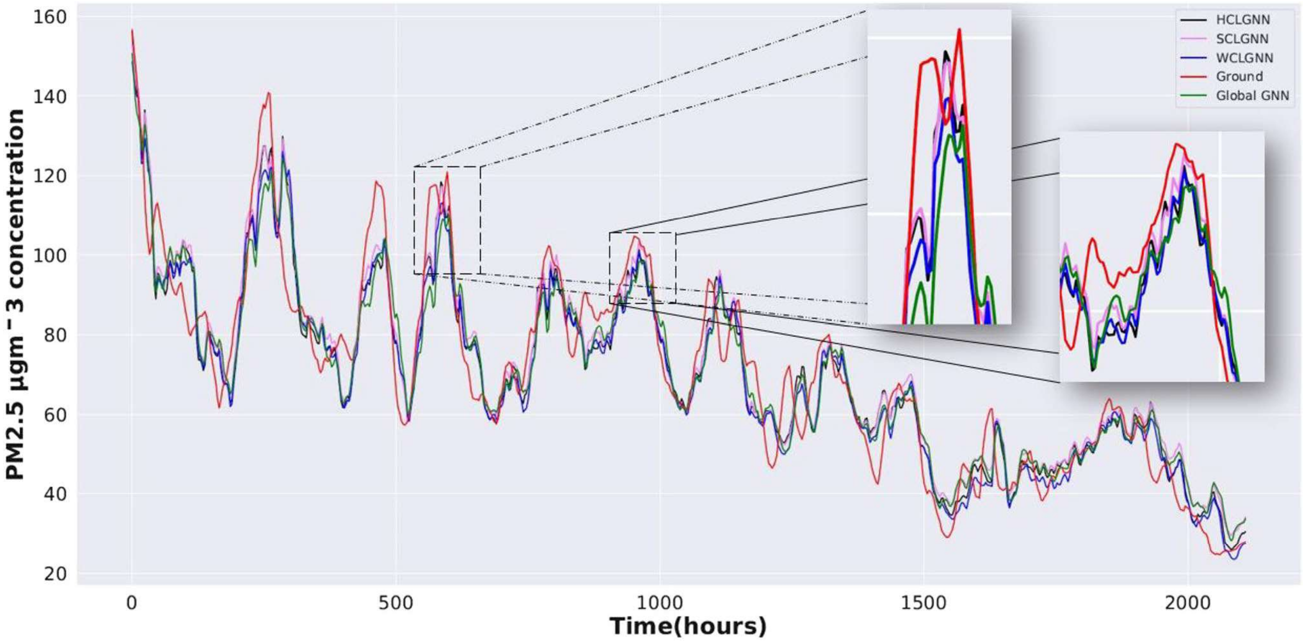


Fig. 10 Spatial average of $PM_{2.5}$ prediction for different models for k=8

$PM_{2.5}$ concentration. For the GMSs in A.P. state WCLGNN, SCLGNN, and HCLGNN shows R^2 value 0.33, 0.34, 0.339 with respect to the R^2 value 0.23 for the Global-GNN model, which is a significant improvement. The low population density and topography of the state may work as a factor for low $PM_{2.5}$ concentration in AP state. On the other hand, Assam, Nagaland GMSs report high $PM_{2.5}$ values, at the same time the proposed models show high accuracy of the forecasts on the test data with R^2 value 0.73 and 0.75. It is noteworthy at this point that, SWFU unit extracts important features related to station-wise meteorological dependency on the $PM_{2.5}$ forecast. Without using this unit, the model performs as a local $PM_{2.5}$-GNN model. For individual clusters, use of SWFU unit provides better R^2 value compared to using local $PM_{2.5}$-GNN model as expected. Furthermore, the meteorological features play an important role in the network. The removal of the weather features decreases the performance of the local models at a significant level.

For 3 different prediction lengths (k= 8, 16, 24), the developed model's prediction capabilities are shown in Fig. 9 in terms of RMSE, MAE and R^2 values. It is observed that, with the increment of prediction steps, the prediction metrics R^2 values decrease thoroughly in this case. Our modeling strategy shows better performance than the Global-GNN for both +24 hours (k=8) and +72 hour (k=24) models. The proposed models' performance in k=24 settings shows a significantly better R^2 values with respect to the Global-GNN model. The individual cluster-wise forecasting accuracies on the test data for each developed CLGNN models (WCLGNN, SCLGNN and HCLGNN) are shown in Table 6 for k=8, Table 7 for $k = 16$ and Table 8 for $k = 24$ respectively. The best accuracy metrics can be observed from the table for each individual clusters along with required training time. The Fig. 10 depicts the model prediction capabilities by taking spatial average for all the GMSs for k=8 as a sample case. The Global GNN model generally underestimates high $PM_{2.5}$ concentrations. The underestimation is relatively less for WCLGNN, SCLGNN and HCLGNN compared to Global GNN model. On the other hand, all these models overestimate for lower concentration values. Cluster-based models promise a less overestimation compared to Global-GNN model. Figure 11 depicts spatially averaged residual error for 3 different cases. It can be seen the residual error of the predictions by proposed models are better with respect to Global GNN model for high prediction variability. Moreover, the spatial RMSE propagation of the proposed models are shown in Fig. 12. It can be visualized, the effect of the proposed models on the RMSE propagation around the IGP region. The proposed models achieve relatively less RMSE around the IGP region and also on the south western part of India with respect to the Global model. For $k = 16$, all the proposed models and $k = 8$, SCLGNN model show lesser RMSE propagation in the south-west part of India (Kerala, Karnataka). Overall, the SCLGNN model shows higher RMSE propagation in the northern part of India and lesser RMSE propagation in

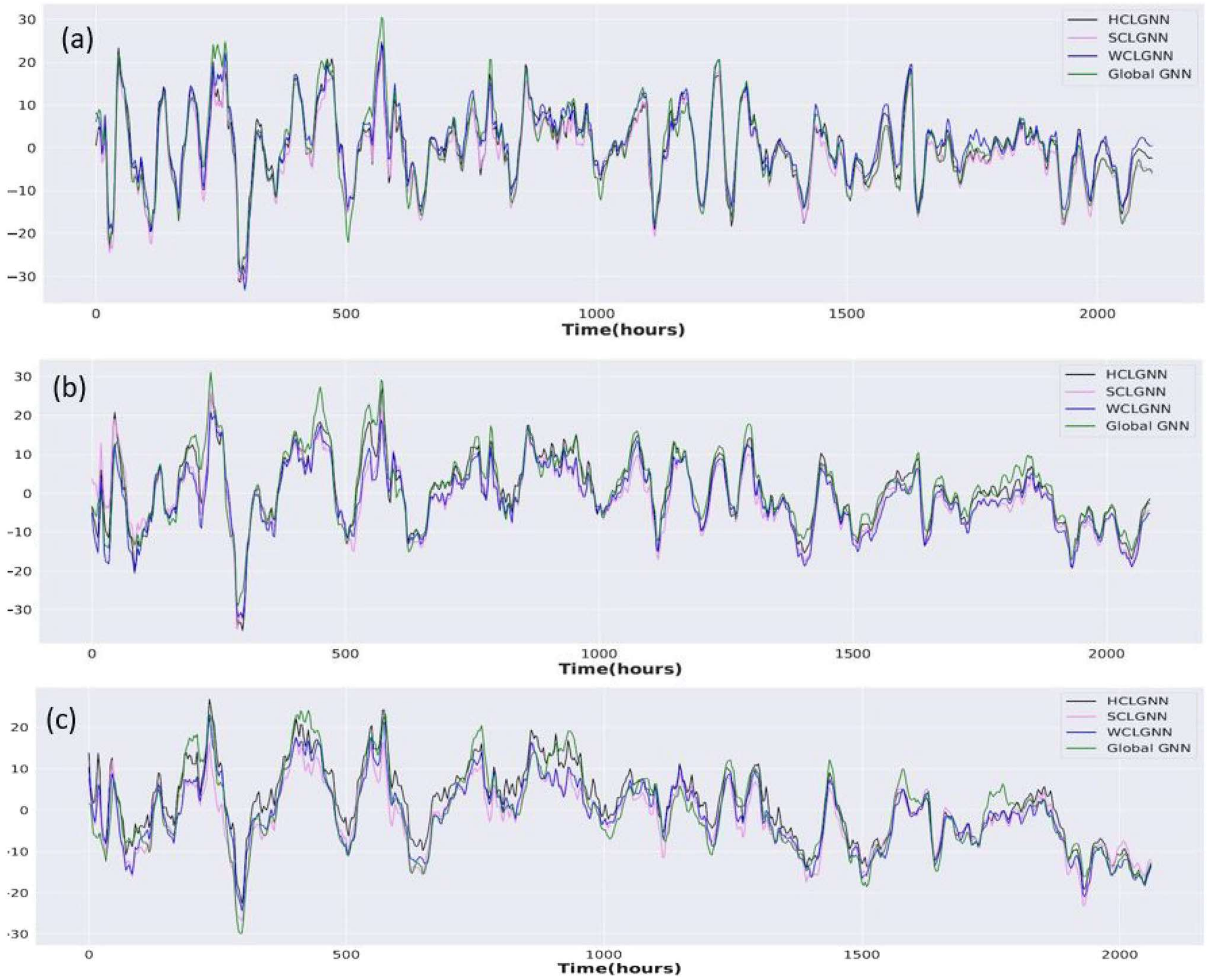


Fig. 11 Spatial average of Residual error for different models on the test data for (**a**) k=8 (24 hour forecasting model) (**b**) k=16 (48 hour), (**c**) k=24 (72 hour forecasting model)

the south-west part of India with respect to other proposed model. In terms of RMSE propagation, WCLGNN model performs better for the IGP region.

3.4 Discussions

Very few studies have considered air pollutant transportation-related effects on the PM$_{2.5}$ concentrations for individual monitoring sites or the adjacent sites in their deep learning based model architectures or methodologies. Several studies (Mengfan et al. 2022; Huang and Kuo 2018) have claimed that the adjacent monitoring sites' meteorological and PM$_{2.5}$ features can help in processing spatial features for air pollution forecasting framework. This study investigates possible mitigation methodologies for different shortcomings of a single GNN based PM$_{2.5}$ modeling over a big graph structure. Firstly, the cluster-based hybrid GNN framework efficiently extracts the relevant adjacent sites feature and reduces complexities by avoiding less significant parameters in the network. For each clusters, the meteorological data mixes with PM$_{2.5}$ data in an efficient manner capturing local PM$_{2.5}$ propagation pattern. Secondly, two different modules were incorporated in the local HGNN models for capturing individual robust prediction capabilities. Among these two modules, the spatio-temporal module uses the advection coefficient (proposed by Wang et al. 2020). This module extracts spatio-temporal feature and concatenates them with feature from the SWFU unit to provide a robust forecast. The pollutant transportation-related features are extracted by the spatio-temporal module of the model. The SWFU unit first only consider the PM$_{2.5}$ and meteorological feature mixing in a station-wise manner, subsequently the mixed feature are

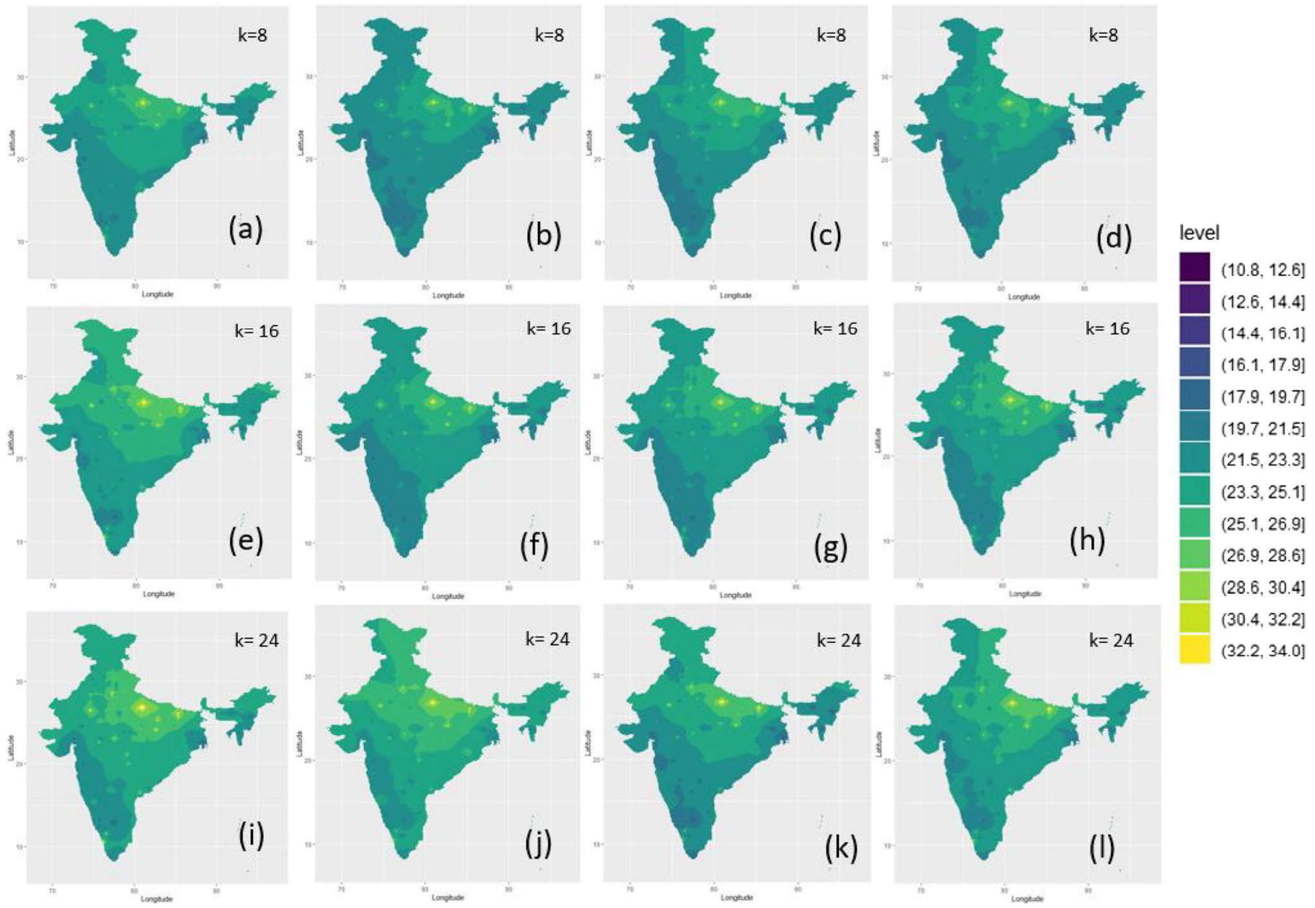


Fig. 12 RMSE propagation of Global GNN and proposed models for on the test data for k=8 (24- hour), 16 (48-hour), 24 (72-hour) models; (**a**), (**e**), (**i**) represents Global-GNN; (**b**), (**f**), (**j**) WCLGNN; (**c**), (**g**), (**k**) SCLGNN and (**d**), (**h**), (**l**) HCLGNN respectively

forwarded to model the temporal dependency through the GRU layer of the module. Using this unit in the model, the accuracy increases for individual clusters.

Mostly, significant meteorological factors such as (boundary layer height (Soh et al. 2018), atmospheric pressure, K-Index) have been overlooked in forecasting pollutants. These important meteorological features are included in this study for the framework on a national level scale for India. Additionally, it is well known that IGP region is one of the most polluted places not only in India

Table 6 Different CLGNN model accuracy over test data with predicted length 8

Algorithm	Cluster	Trainloss	Testloss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR	Parameters	R^2	Time (min)
Weighted k-means	1(123)	0.370	0.180	20.09	16.71	0.750	0.877	0.163	422756	0.690	165
	2(48)	0.260	0.121	17.41	14.46	0.710	0.817	0.156	255956	0.698	89
	3(34)	0.297	0.161	19.32	16.11	0.765	0.875	0.141	241018	0.670	100
	4(53)	0.328	0.138	18.29	15.31	0.668	0.825	0.223	262526	0.664	65
Spectral	1(46)	0.353	0.199	22.25	18.98	0.723	0.946	0.246	253510	0.635	60
	2(92)	0.394	0.187	21.05	17.71	0.752	0.901	0.180	336080	0.679	92
	3(54)	0.271	0.121	17.51	14.62	0.710	0.842	0.182	263918	0.698	65
	4(66)	0.327	0.129	16.92	13.97	0.686	0.802	0.174	279605	0.696	80
Hierarchical	1(166)	0.368	0.179	20.20	16.88	0.744	0.901	0.190	84350	0.681	543
	4(92)	0.291	0.128	17.79	14.80	0.689	0.821	0.189	336080	0.675	139

Table 7 Different CLGNN model accuracy over test data with predicted length 16

Algorithm	Cluster	Trainloss	Testloss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR	Parameters	R^2	Time (min)
Weighted k-means	1(123)	0.398	0.205	22.68	18.42	0.724	0.875	0.192	422756	0.633	388
	2(48)	0.283	0.136	19.66	16.23	0.692	0.874	0.232	255956	0.648	96
	3(34)	0.298	0.178	21.61	17.72	0.748	0.931	0.209	241018	0.618	122
	4(53)	0.331	0.147	19.73	16.16	0.655	0.836	0.248	262526	0.628	102
Spectral	1(46)	0.367	0.207	22.79	18.61	0.738	0.912	0.206	253510	0.609	94
	2(92)	0.413	0.204	23.04	18.67	0.724	0.859	0.178	336080	0.634	144
	3(54)	0.302	0.141	20.11	16.52	0.687	0.863	0.228	263918	0.631	104
	4(66)	0.350	0.152	19.98	16.41	0.658	0.856	0.260	279605	0.629	124
Hierarchical	1(166)	0.394	0.203	22.56	18.29	0.728	0.894	0.204	584350	0.622	819
	4(92)	0.310	0.140	19.35	15.61	0.671	0.797	0.190	336080	0.633	211

Table 8 Different CLGNN model accuracy over test data with predicted length 24

Algorithm	Cluster	Trainloss	Testloss	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	CSI	POD	FAR	Parameters	R^2	Time (min)
Weighted k-means	1(123)	0.417	0.220	24.51	19.68	0.708	0.899	0.230	422756	0.598	520
	2(48)	0.280	0.143	20.50	16.58	0.685	0.869	0.236	255956	0.621	128
	3(34)	0.314	0.177	21.87	17.37	0.741	0.886	0.181	241018	0.612	258
	4(53)	0.333	0.148	19.89	15.88	0.639	0.765	0.204	262526	0.619	137
Spectral	1(46)	0.373	0.216	23.40	18.65	0.735	0.878	0.181	253510	0.585	322
	2(92)	0.438	0.226	24.76	19.90	0.723	0.905	0.218	336080	0.591	283
	3(54)	0.305	0.149	20.68	16.61	0.683	0.853	0.226	263918	0.603	134
	4(66)	0.348	0.158	20.70	16.84	0.649	0.851	0.268	279605	0.608	166
Hierarchical	1(166)	0.392	0.207	23.16	18.34	0.723	0.871	0.191	584350	0.607	1127
	4(92)	0.312	0.150	20.56	16.49	0.657	0.818	0.230	336080	0.601	284

but in the world. This cluster-based methodology is suitable for capturing the long-range transport of the air pollutants in the IGP region. The proposed models provide significantly better performance in IGP region with respect to the baseline models. Literatures revealed that wind speed is an important feature responsible for transport of $PM_{2.5}$ (Zeydan and Wang 2019). Few monitoring stations in the western Madhya Pradesh state, Delhi, Tripura, Vijaywada (A.P.) reflect high mean wind speed over the year. The proposed models have shown high R^2 values on the GMSs (with high mean wind speed) with respect to the Global-GNN model. Among the proposed GNN models, the WCLGNN and HCLGNN models work close to the same precision on the IGP region, but HCLGNN comes with higher training time. In terms of RMSE propagation, 24 hour forecasting-SCLGNN model performs better in the south-west India. The SCLGNN model comes with the least training time but provide almost similar accuracy for all the prediction horizon studied in the work. The WCLGNN model metrics show a very less improvement with high training time compared to SCLGNN for all the considered predicted lengths (k). Also, WCLGNN model requires a high number of parameters to train with respect to other models. As HCLGNN has only two clusters, it necessitates a lengthy so this model demands a high training period.

In general, use of graph neural network based models for air quality forecasting purpose is limited (Qi et al. 2019; Huang et al. 2021; Wang et al. 2020 etc.) The Qi et al. developed GC-LSTM model over 76 monitoring stations with +72 hour, +48 hour, +24 hour forecasting capabilities with forecasting accuracy metrics MAE 24.21, 23.66, 23.10 $\mu g/m^3$ and RMSE 38.83 $\mu g/m^3$, 37.32 $\mu g/m^3$, 34.19 $\mu g/m^3$ respectively. The coverage area of the monitoring stations are located in the North-Eastern part of China. Whereas, our framework provides forecasting accuracy metrics for best +72 hour, +48 hour, +24 hour forecasting capacities in terms of MAE are 15.91 $\mu g/m^3$, 17.29 $\mu g/m^3$, 17.64 $\mu g/m^3$. Comparing the air pollution profile of India with respect to

China, differs on the basis of sociological, demographical and geography of the countries. With a least comparative point of view, our Cluster-based GNN models show superiority considering the underlying Graph structure and scale of the study region. Apart from that, our models show better performance than $PM_{2.5}$-GNN as well.

4 Conclusion

The increasing availability of real-time observations and computing resources motivates us to develop a deep learning model for air pollution forecasting. In this study a cluster-based ensembled GNN methodology was proposed for air quality forecasting all over India. Our modeling explores the integration of Graph Neural Network at a local scale. Its implementation is carried out in a zone-wise manner to capture local spatial features based on local meteorological conditions. With acceptable prediction capabilities, the proposed modeling approach is verified to be an effective method to forecast ground $PM_{2.5}$ concentrations based on historical observations on a national level scale. With shorter training time, the developed cluster-based models outperform the state-of-the-art PM2.5-GNN model. For +24 hour forecasting, among various prediction horizons, the CLGNN models achieve the maximum accuracy in terms of RMSE, MAE, and R^2 value. Also, WCLGNN model performs better than other proposed models on IGP region.

The use of finer spatial resolution of meteorological variables data, can improve the forecasting capability of the proposed models in a significant level. Many GMS are located in high numbers at the metropolitan cities of India, so with the use of ERA5 dataset in 30 km resolution, slight meteorological changes can not be extracted by the proposed framework. In this methodology, one can further develop robust local models respecting the local population density, anthropological causes, topography and the atmospheric conditions for $PM_{2.5}$ concentrations forecasting.

Acknowledgements The Authors are thankful to Central Pollution Control Board of India (https://cpcb.nic.in/) for $PM_{2.5}$ data and Copernicus Climate Change Service (C3S) for ERA5 dataset.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

Conflicts of interest The authors declare no conflict of interest.

References

Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal, R.S., Brauer, M., Cohen, A.J., Dandona, L.: The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. Lancet Planet. Health **3**(1), e26–e39 (2019)

Becerra-Rico, J., Aceves-Fernández, M.A., Esquivel-Escalante, K., Pedraza-Ortega, J.C.: Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks. Earth Sci. Inf. **13**(3), 821–834 (2020)

Beddows, D.C., Dall'Osto, M., Harrison, R.M.: Cluster analysis of rural, urban, and curbside atmospheric particle size data. Environ. Sci. Technol. **43**(13), 4694–4700 (2009)

Byun, D., Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system (2006)

Chakma, A., Vizena, B., Cao, T., Lin, J., & Zhang, J.: Image-based air quality analysis using deep convolutional neural network. In: IEEE International Conference on Image Processing (ICIP), pp. 3949–3952. IEEE (2017)

Chung, J., Gulcehre, C., Cho, K., Bengio, Y: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555 (2014)

Copernicus Climate Change Service (C3S) ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS), (2017). https://cds.climate.copernicus.eu/cdsapp#!/home. Accessed 31 Sept 2010

CPCB Central Pollution Control Board: Air quality in Delhi (1989–2000), National Ambient Air Quality Monitoring Series-NAAQMS/17/2000-2001. Parivesh Bhawan, Delhi, India (2001)

Davies, D. L., Bouldin, D. W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. (2), 224–227 (1979)

Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems, 29 (2016)

Forkel, R., Werhahn, J., Hansen, A.B., McKeen, S., Peckham, S., Grell, G., Suppan, P.: Effect of aerosol-radiation feedback on regional air quality-A case study with WRF/Chem. Atmos. Environ. **53**, 202–211 (2012)

Galmarini, S., Kioutsioukis, I., Solazzo, E., Alyuz, U., Balzarini, A., Bellasio, R., Unal, A.: Two-scale multi-model ensemble: is a hybrid ensemble of opportunity telling us more? Atmos. Chem. Phys. **18**(12), 8727–8744 (2018)

Gogikar, P., Tyagi, B., Gorai, A.K.: Seasonal prediction of particulate matter over the steel city of India using neural network models. Model. Earth Syst. Environ. **5**(1), 227–243 (2019)

Goodfellow, I., Bengio, Y., Courville, A: Deep learning. In: MIT Press (2016)

Grivas, G., Chaloulakou, A.: An overview of the PM_{10} pollution problem, in the Metropolitan Area of Athens, Greece. Assessment of controlling factors and potential impact of long range transport. Sci. Total Environ. **389**(1), 165–177 (2008)

Hu, J., Wang, Y., Ying, Q., Zhang, H.: Spatial and temporal variability of $PM_{2.5}$ and PM_{10} over the North China Plain and the Yangtze River Delta, China. Atmos. Environ. **95**, 598–609 (2014)

Hu, Y., You, H., Wang, Z., Wang, Z., Zhou, E., Gao, Y.: Graph-MLP: node classification without message passing in graph. arXiv:2106.04051 (2021)

Huang, C.J., Kuo, P.H.: A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors **18**(7), 2220 (2018)

Huang, J.Z., Ng, M.K., Rong, H., Li, Z.: Automated variable weighting in k-means type clustering. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 657–668 (2005)

Huang, Y., Ying, J.J.C., Tseng, V.S.: Spatio-attention embedded recurrent neural network for air quality prediction. Knowl.-Based Syst. **233**, 107416 (2021)

Kang, D., Mathur, R., Schere, K., Yu, S., Eder, B.: New categorical metrics for air quality model evaluation. J. Appl. Meteorol. Climatol. **46**(4), 549–555 (2007)

Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Sachdeva, S.: Evaluation of different machine learning approaches to forecasting PM2.5 mass concentrations. Aerosol Air Qual. Res. **19**(6), 1400–1410 (2019)

Lal, B., Tripathy, S.S.: Prediction of dust concentration in open cast coal mine using artificial neural network. Atmos. Pollut. Res. **3**(2), 211–218 (2012)

Li, Y., Lang, J., Ji, L., Zhong, J., Wang, Z., Guo, Y., He, S.: Weather forecasting using ensemble of spatial-temporal attention network and multi-layer perceptron. Asia-Pac. J. Atmos. Sci. **57**(3), 533–546 (2021)

Lin, Y., Mago, N., Gao, Y., Li, Y., Chiang, Y. Y., Shahabi, C., Ambite, J. L.: Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In: Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 359–68 (2018)

Ma, J., Li, K., Han, Y., Du, P., Yang, J.: Image-based pm2.5 estimation and its application on depth estimation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1857–1861. IEEE (2018)

Mengfan, T., Siwei, L., Lechao, D., Senlin, H.: Including the feature of appropriate adjacent sites improves the $PM_{2.5}$ concentration prediction with long short-term memory neural network model. Sustain. Cities Soc. **74**, 103427 (2022)

Miao, Y., Li, J., Miao, S., Che, H., Wang, Y., Zhang, X., Liu, S.: Interaction between planetary boundary layer and $PM_{2.5}$ pollution in megacities in China: a review. Curr. Pollut. Rep. **5**(4), 261–271 (2019)

Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: Analysis and an algorithm. Adv. Neural Inf. Proces. Syst. **14**, 849–856 (2001)

Pant, P., Lal, R.M., Guttikunda, S.K., Russell, A.G., Nagpure, A.S., Ramaswami, A., Peltier, R.E.: Monitoring particulate matter in India: recent trends and future outlook. Air Qual. Atmos. Health **12**(1), 45–58 (2019)

Qi, Y., Li, Q., Karimian, H., Liu, D.: A hybrid model for spatiotemporal forecasting of $PM_{2.5}$ based on graph convolutional neural network and long short-term memory. Sci. Total Environ. **664**, 1–10 (2019)

Rao, N.D., Kiesewetter, G., Min, J., Pachauri, S., Wagner, F.: Household contributions to and impacts from air pollution in India. Nat. Sustain. **4**(10), 859–867 (2021)

Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)

Rumelhart, D. E., Hinton, G. E., Williams, R. J.: Learning internal representations by error propagation. In: California Univ San Diego La Jolla Inst for Cognitive Science (1985)

Senior, A., Heigold, G., Ranzato, M. A., Yang, K.: An empirical study of learning rates in deep neural networks for speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal (2013)

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM National Conference (1968)

Soh, P.W., Chang, J.W., Huang, J.W.: Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access **6**, 38186–38199 (2018)

Spindler, G., Brüggemann, E., Gnauk, T., Grüner, A., Müller, K., Herrmann, H.: A four-year size-segregated characterization study of particles PM_{10}, $PM_{2.5}$ and PM_1 depending on air mass origin at Melpitz. Atmos. Environ. **44**(2), 164–172 (2010)

Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Kerschbaumer, A., et al.: A model inter-comparison study focussing on episodes with elevated PM10 concentrations. Atmos. Environ. **42**(19), 4567–4588 (2018)

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J., Favez, O., et al.: Meteorology-driven variability of air pollution (PM_1) revealed with explainable machine learning. Atmos. Chem. Phys. **21**(5), 3919–3948 (2021)

Tai, A.P., Mickley, L.J., Jacob, D.J.: Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change. Atmos. Environ. **44**(2), 3976–3984 (2010)

Tiwari, A., Gupta, R., Chandra, R.: Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown. arXiv:2102.10551 (2021)

Tong, W., Li, L., Zhou, X., Hamilton, A., Zhang, K.: Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN. Air Qual. Atmos. Health **12**(4), 411–423 (2019)

Van der Hoven, I.: Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. J. Atmos. Sci. **14**(2), 160–164 (1957)

Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., Gao, F.: Pm2.5-gnn: A domain knowledge enhanced graph neural network for pm2.5 forecasting. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp. 163–166 (2020)

Wu, S., Li, G., Chen, F., Shi, L.: Training and inference with integers in deep neural networks. arXiv:1802.04680 (2018)

Xing, Y., Xu, F., Shi, Y.H., M.H., Lian, Y.X.: The impact of $PM_{2.5}$ on the human respiratory system. J. Thorac. Dis. **8**(1), E69 (2016)

Xu, X., Tong, T., Zhang, W., Meng, L.: Fine-grained prediction of $PM_{2.5}$ concentration based on multisource data and deep learning. Atmos. Pollut. Res. **11**(10), 1728–1737 (2020)

Zeydan, Ö, Wang, Y.: Using MODIS derived aerosol optical depth to estimate ground-level $PM_{2.5}$ concentrations over Turkey. Atmos. Pollut. Res. **10**(5), 1565–1576 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

View publication stats